In this manuscript, the authors simulate frozen ground and streamflow across a pan-Arctic domain. They simulate increased runoff, higher proportions of subsurface flow and a loss of permafrost. A notable finding is a regional shift in runoff towards more northerly locations, which have higher amounts of soil carbon. This leads to the conclusion that there could be enhanced carbon fluxes to the Arctic Ocean. The paper is well written and presented. I have several suggestions, which are embedded in the attached document, so I will only summarize in these comments.

The only major concern I have is the validation of the representation of frozen ground. It is unclear how the authors represent discontinuous permafrost within a model grid. Furthermore, I am not sure how the simulated ground thermal state is compared to observations. I struggle with using permafrost class to validate simulated active layer thickness. A better model validation approach is necessary.

The authors appreciate the reviewer's assessment and comments. There are two components to clarify with respect to the validation of frozen ground.

Regarding permafrost classification, permafrost class from the International Permafrost Association (IPA) dataset (Brown et al., 1997) was presented simply to illustrate the spatial pattern in permafrost distribution for the benefit of a broad readership. It was not used to validate simulated active-layer thickness (ALT). For that the authors leveraged a new dataset, (TPDC), which was produced using a machine-learning approach that incorporated in situ data of permafrost thermal state and ALT (Ran et al., 2022). Results of that comparison are documented in figure 1d of the submitted manuscript. While the TPDC data set is state-of-art, there are inherent uncertainties in its spatial representation of ALT, as discussed by Ran et al. (2022), and as discussed in the reply to review #1. In essence, Ran et al. specifically stated that the distribution coming out of the machine-learning method is, in their estimation, constrained. Thus the validation results shown in figure 1d are excellent given the known challenges in simulating soil temperatures.

Brown, J., O.J. Ferrians, Jr., J.A. Heginbottom, and E.S. Melnikov, eds. 1997. Circum-Arctic map of permafrost and ground-ice conditions. Washington, DC: U.S. Geological Survey in Cooperation with the Circum-Pacific Council for Energy and Mineral Resources. Circum-Pacific Map Series CP-45, scale 1:10,000,000, 1 sheet.

Ran, Y., Li, X., Cheng, G., Che, J., Aalto, J., Karjalainen, O., Hjort, J., Luoto, M., Jin, H., Obu, J. and Hori, M., 2022. New high-resolution estimates of the permafrost thermal state and hydrothermal conditions over the Northern Hemisphere. Earth system science data, 14(2), pp.865-884.

Regarding discontinuous permafrost, there is no dealing with that, per se. The spatial domain is discretized into an array of grid cells. To determine permafrost state for a grid cell, soil temperatures that vary with depth (23 layers to 60 meters) are examined. A grid cell is deemed to be permafrost or seasonally frozen (non-permafrost) based on the grid cell temperature profile. In the case where soil temperatures are well simulated, one can assume that there is discontinuous permafrost in regions where many grid cells classified as permafrost interface with many grid cells classified as seasonally frozen. Similar to other studies using a land surface model, permafrost state is a binary classification. No subgrid variability is implemented. The authors are confident that no systematic bias exists with this approach. That is, in areas where a majority of cells are classified as permafrost, a significant amount of terrain would, in the real world, contain permafrost, with no permafrost on, for example, south facing slopes. Likewise, in areas where the model simulation points to no permafrost, much of the terrain would have no frozen ground to depth, with patchy areas of permafrost on, for example, north facing slopes. Lines 179-186 in the submitted manuscript describe how the presence of permafrost is determined. More relevant information on simulation ALT and associated challenges is articulated in response to reviewer RC1.

Could I suggest the authors present panels of differences (i.e., GLEAN subtract PWBM) in Figures 1 and 2 to show how different the model is from observations? This would reveal where uncertainty is highest. This is important because one of the main takeaways is the importance of regional differences in responses to warming and the impact this has on freshwater fluxes to the Arctic Ocean.

Further to this point, the conclusion that the model performs well is based on an assessment of the model's performance over the entire domain. This is why I am suggesting those extra panels in Figures 1 and 2 as they will help in assessment of uncertainty across sub-domains.

Finally, I like the paper and what the authors are trying to accomplish. I am hoping that addressing these suggestions will make the conclusions in the final version of the paper a bit more defensible, and impactful. On this note, I think the conclusions could be even punchier. Please see my suggestions in the marked-up version.

Thanks for the chance to review the paper.

Difference at grid cell level have been computed. In light of this comment the authors propose to create map panels, add them, and describe differences in a paragraph to be added to section 3 ‘Model Validation’.

Specific comments:

Lines 115-119: The authors are willing to add a statement to that effect.

Line 122: Appreciate the suggestion.

Line 142: The authors agree, and can do this.

Line 190: Language corrections noted.

Line 195: Only simulated ALT was examined from all three forcings. This in light of the importance of simulated ALT. The authors have clarified the excellent performance of the model simulation of ALT in responses to reviewer #1. The manuscript describes all the remaining results that consistently leverage just the W5E5-based estimates for model validation.

Line 216: The phrase “business as usual” is appropriate.

Line 263: The authors are willing to create the map panels and add them.

Comment on figure 1: Permafrost class addressed in major comments section above. A difference map for ALT to be added.

Line 277: Respectfully, differences of less than 13% is an excellent result for numerical model simulation of Arctic river discharge.

Line 293: Map panels of difference in figures 1 and 2 to be added.

Line 313: The authors disagree with the reviewers assertion. No modeling studies published in recent years suggest that rainfall in Arctic regions will not increase.

Line 328: There is a large body of prior research based on results using this model. The authors believe that addition of extensive model validation discussion would be awkward given the length and breadth of the submitted manuscript. The authors are willing to add map panels of differences shown in figures 1 and 2.

Line 328: The nature of grid cell permafrost classification is articulated above.

Line 353: The authors propose addition of a paragraph in the model validation section which will focus on regional uncertainties based on map graphics in figures 1 and 2.

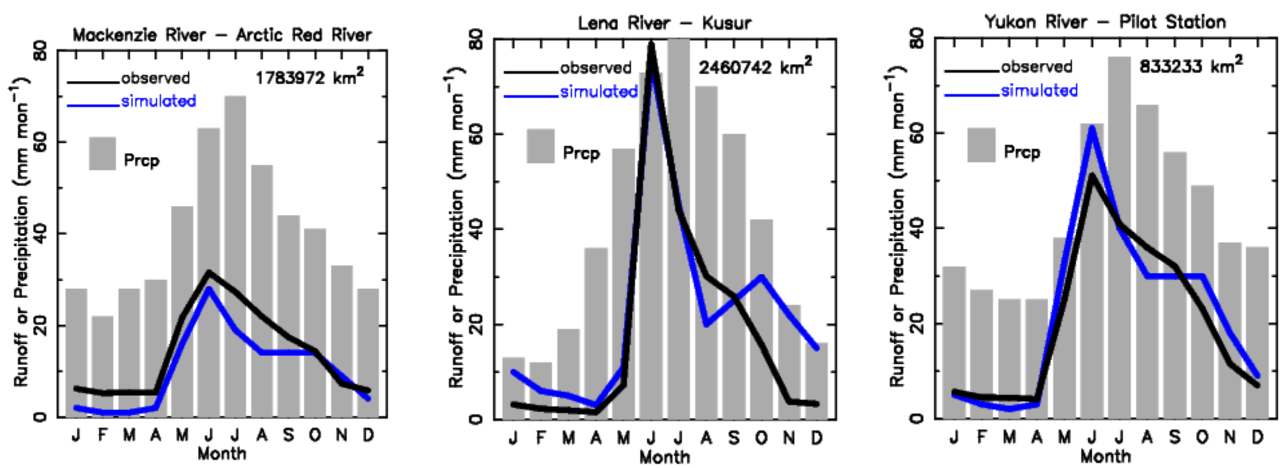
Line 356: Yes. The close correspondence between P-ET and runoff support the statement that changes (increases and decreases) in net precipitation are the main driver of runoff changes.

Line 373: Runoff is defined at lines 157-160. The authors are willing to add another sentence.

Caption figure 6: Change to be made on revision.

Line 404: This is an artifact of the interpolation of IPSL climate fields.

Line 410: Respectfully, there is evidence, though not explicitly demonstrated with a figure graphic in the present manuscript. The authors feel that additional model validation of this nature is far beyond the scope of the present study. It would require assembly of monthly climatologies of river discharge for major rivers, translated to runoff per unit depth across contributing watersheds, and subsequent assessment against model simulations. Rather, we show here prior assessments made during an earlier study. The authors believe that the model simulations are robust enough to support the analysis and statements characterized at line 410.



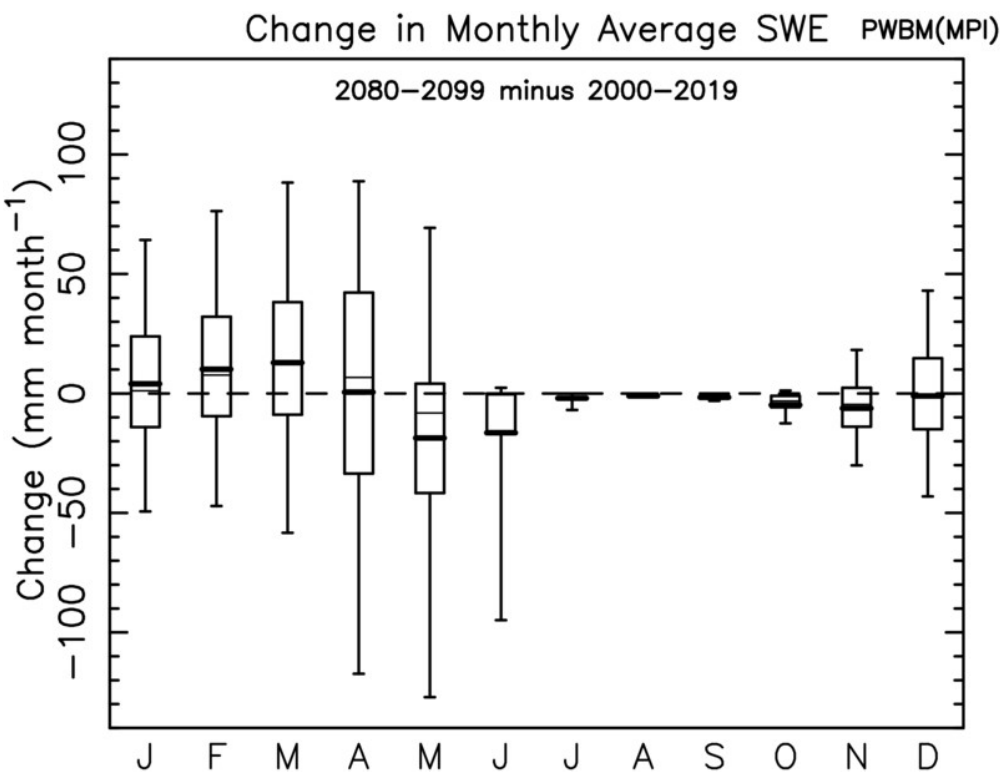
Line 415: Will be moved to Methods on revision.

Line 421: Statement is correct; the mean change in June and July are not significant due to the broad distribution of change across the domain as shown by the boxplots. A phrase will be added to the sentence to clarify this point.

Line 465: Independent clause is based on the amount of net precipitation change in the two simulations described. De-watering not quantified in the present study. Can reword to state that de-watering of permafrost plays a much smaller role, citing McClelland et al. (2004).

McClelland, J. W., Holmes, R. M., Peterson, B. J., and Stieglitz, M. (2004), Increasing river discharge in the Eurasian Arctic: Consideration of dams, permafrost thaw, and fires as potential agents of change, J. Geophys. Res., 109, D18102, doi:10.1029/2004JD004583.

Line 492-494: Yes, the statement is based entirely on model simulations and examination of associated outputs. The monthly distributions were examined. Moreover, there is no increase in model simulated SWE in autumn. Figures not shown in manuscript given the large number of existing graphics.



Line 506: Agreed with reviewer comment. Clause regarding other studies will be added on revision.

Line 539: Increased soil thaw in the simulations in areas rich with soil carbon support the second part, and is consistent with studies based on river sampling.

Line 546: Agree that details of approximations for PET would overwhelm readers. The descriptions in this paper are more succinct given that three more detailed papers have been published in recent years. A model simulating AET is too computationally expensive for a study of this type, and, moreover, forcing data are lacking. Using a function like Penman-Monteith would introduce much more uncertainty. A cost benefit analysis suggests that using the Haman function is appropriate. After all, the study is not focused on accuracy of land-atmosphere water fluxes. That said, mention of Hamon function will be added to Methods section.

Line 575: The authors appreciate the comment. The authors opine that uncertainties have been articulated where appropriate. The PWBM is suitably physically scaled for a study of this type. It is not a land-surface model of the type used in coupled climate model simulations. Extensive discussion of model uncertainties would be a different paper altogether. The present manuscript is not focused on a rigorous assessment of validation exercises with the PWBM. The present results build on a rich history of prior studies, each of which contain model validations suitable for the study objectives and goals. Detailed validation at regional scale is beyond the scope of the present study which seeks to gain insights into future trajectories. Moreover, all of the results, namely, that runoff will shift northward, and to more subsurface flow and toward later in autumn, are consistent with recent studies that leverage river sampling and in situ ground measurements. In revision the authors will add statements based on difference maps arising from fields shown in figures 1 and 2.